

法政大学学術機関リポジトリ

HOSEI UNIVERSITY REPOSITORY

騒音下における拡声器のための実時間音声加工システムの開発

著者	河野 日向子
出版者	法政大学大学院理工学研究科
雑誌名	法政大学大学院紀要．理工学・工学研究科編
巻	60
ページ	1-6
発行年	2019-03-31
URL	http://doi.org/10.15002/00022068

騒音下における拡声器のための 実時間音声加工システムの開発

DEVELOPMENT OF REAL-TIME VOICE PROCESSING SYSTEM FOR THE LOUDSPEAKERS USED UNDER NOISY ENVIRONMENTS

河野日向子

Hinako KAWANO

指導教員 鈴木郁

法政大学大学院理工学研究科システム理工学専攻修士課程

It is well known that elderly people may have difficulties to listen speech spread by public-address systems, especially under noisy environments. The aim of this study is to develop a real-time voice processing system used with a public address-system used under noisy environments. The processing system is intended to improve the intelligibility of the public address system to assist the elderly, without increasing the sound pressure level of the voice. An experiment to confirm the effectiveness of the system was conducted on two elderly persons. In the experiment, they were asked to write down compressed or uncompressed words under noises. The difference in percent correct answers between compressed voice and uncompressed one was not statistically significant, however.

Key Words : voice, noise, Senior citizens, real-time, loudspeaker

1. 研究の目的

日常生活において、周囲騒音が拡声器により拡声される音声の聞き取りやすさを低下させることがある。特に、高齢者は聴覚機能が低下するため、若年者よりもその影響を受けやすい。この最も単純な解決策は、音声の音圧レベルを大きくすることであるが、若年者に不快感を与えてしまうという問題がある。これを回避するために、過去に、多帯域振幅圧縮処理を利用した音声加工システムの研究がされており、騒音に感応して音声を加工することの有効性は認められている¹⁾。しかし、実時間処理でないため実環境で使用することはできない。そこで本

研究では、実時間で音声聞き取りやすくなるように加工をするシステムの開発を目的とする。一例として、駅構内において本システムを使用した際のシステムの仕組みを簡略化し、図1に示す。このシステムは図のように、マイクロホンから取得した騒音に感応して音声を加工し、加工した音声をスピーカーへと出力する。

2. 多帯域振幅圧縮

振幅圧縮とは、入力に対して出力の音圧レベルの変化を減じる信号処理のことである。ここで、未圧縮音声と圧縮音声の入出力値の音圧レベルをそれぞれ図2の(a)、(b)に示す。(a)は入力と出力のダイナミックレンジが等しいのに比べて、(b)は入力に対して出力のダイナミッ

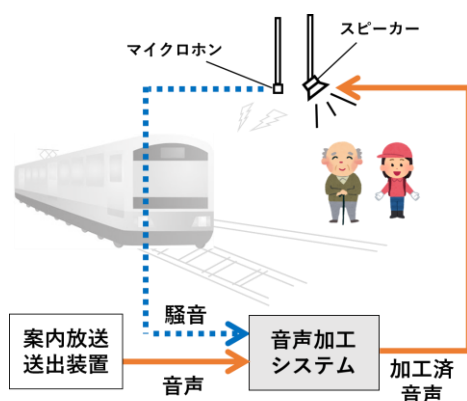


図1 音声加工システムの仕組み

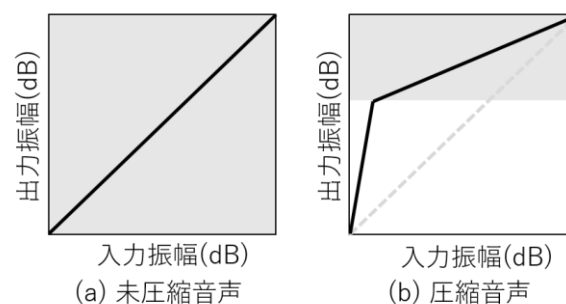


図2 入出力値の音圧レベル

クレンジが狭い。この振幅圧縮を多帯域に分けて行う処理を、多帯域振幅圧縮処理²⁾という。また、本システムでは入力音圧レベルの最大値と出力音圧レベルの最大値を合わせているため、振幅圧縮された音声信号の振幅は拡大される。これは最後に全体の音圧レベルを補正することで不変に保つことができる。

3. 方法

本研究で作成するシステムは、サウンド入出力機能を備えた一般的なパーソナルコンピュータと、その上で実行されるLabVIEWで記述されたプログラムにより構成されている。

4. 音声加工システム

(1) システムの概要

このシステムは、周囲騒音の分析結果に応じて各周波数帯域の圧縮の程度（圧縮レベル）を自動で決定し、その圧縮レベルに基づいて多帯域振幅圧縮処理を行うものである。図3にモデル化した音声加工システムの構成を示す。本システムは大きく分けると、多帯域振幅圧縮処理システムと圧縮レベル算出システムから構成されている。

(2) 多帯域振幅圧縮処理システム

多帯域振幅圧縮システムは、目的音声の音圧レベルと、後述する圧縮レベル算出システムから得られる圧縮レベルをもとに多帯域振幅圧縮処理をして、圧縮された音声を出力するものである。本稿では多帯域振幅圧縮の詳細については省略するが、先行研究¹⁾を参考にした。

(3) 圧縮レベル算出システム

圧縮レベル算出システムは、周囲騒音の各帯域の音圧レベルから、帯域毎の圧縮レベルを多帯域振幅圧縮処理システムへ出力するものである。このシステムでは、騒音の各帯域の音圧レベルを入力すると、帯域毎の適切な圧縮レベルを出力するような関数が必要である。適切な圧縮レベルとは、ある騒音下で音声の聞き取りやすくなるようにヒトが手で定めた圧縮レベルを理想としている。このような関数を理論的に作ることは容易ではないため、ニューラルネットワークを使用する。

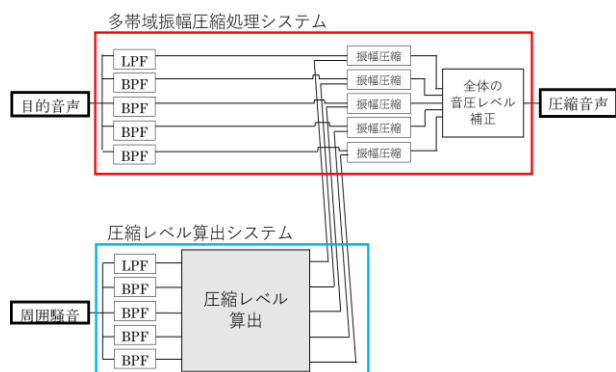


図3 音声加工システムのモデル

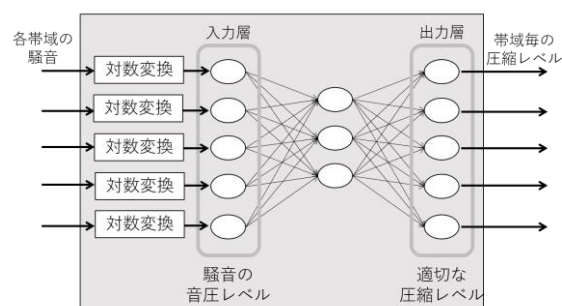


図4 圧縮レベル算出システムの構成

今回、使用したニューラルネットワークは入力層5素子、中間層3素子、出力層5素子で構成し、バックプロパゲーション法により、騒音の各帯域の音圧レベルに対して帯域毎の適切な圧縮レベルを出力するように予め学習させておく。実際には、入力層と中間層に、入力に関係なく常に1を出力するバイアスニューロンを設けている。また、0から1までのあらゆる中間的な値が出力されることを望んでいるため、本研究では通常のシグモイド関数ではなく、

$$F(x) = \frac{1}{1 + e^{-0.5x}} \quad (1)$$

を用いる。このシステムの構成を図4に示す。

5. 教師データ作成実験

(1) 実験の概要

前述のように、騒音の音圧レベルに応じて適切な圧縮レベルを出力するために、ニューラルネットワークを使用する。そのため、様々な騒音に応じて、目的音声のどの帯域をどの程度圧縮すると、音声の聞き取りやすさが改善されるのかを調べる実験を行い、その結果をニューラルネットワークに学習させる必要がある。

実験では、ある騒音下で複数の単語を被験者に呈示し、すべての単語を聞き取ることができるようになるまで、試作した多帯域振幅圧縮処理システムで圧縮レベルを調節させる。さらに、複数の騒音下において同様に実験を行い、それぞれの騒音下において最適な圧縮レベルを決定する。ここで、被験者が音声を聞き取りやすいと判断した時の、騒音毎の各帯域の圧縮レベルを教師データ、騒音毎の各帯域の音圧レベルをトレーニングデータとして保存する。このデータを収集するための被験者実験のことを、以下では教師データ作成実験と呼ぶ。

(2) 使用音源

a) 呈示する音声

目的音声として、補聴器適合評価用CDの“TY-89”^[3]の2音節の単語50種類と3音節の単語50種類、計100種類を使用する。教師データ作成実験では、本来なら高齢者を被験者とするべきであるが、実験時は狭い空間

の中で長時間の作業を行わせなければならないため、高齢者を被験者とすることは困難である。そこで、呈示する音声に、高齢者の加齢に伴う聴覚機能の低下を模擬⁴⁾する加工を施すことで、若年者が実験をした時の結果を、高齢者が実験をした時と同じ結果として扱うことにした。

加齢に伴う聴覚機能の低下は、周波数分解能の低下、周波数分解能の低下、時間分解能の低下の3つの要因が挙げられる。実験で音声を呈示する際は、図5に示す過程で高齢者の聴覚機能低下の模擬をする処理を行う。まず、多帯域振幅圧縮をする前の音声には予め、ホルマント帯域幅を4倍に広げる加工を施す。これにより、周波数分解能の低下を模擬している。次に、多帯域振幅圧縮したあとの音声と騒音に、図6のようなフィルターを用いて減衰処理を施すことで、純音聴力の低下を模擬する。なお、本来であれば時間分解能低下についても模擬を行うべきであるが、今回のシステムがその補償するような加工を行わないことから実験の結果に与える影響は少ないと考え、模擬はしないことにする。

b) 呈示する騒音

実験で使用する騒音は、本システムの将来的な利用環境の代表として鉄道の駅構内などを想定している為、実際の駅構内及び駅周辺で発生する騒音の中から構内放送を覆い隠すと思われる騒音を収集したものを使用する。しかし、この実験は被験者に騒音下で単語を呈示し、聞き取り試験を行うというものであり、この実験で時々刻々変化している実際の騒音を使用すると、結果にばらつきが出てしまい、騒音毎の最適な音声の圧縮程度を決めることは困難である。そこで、実際の騒音をそのまま用いるのではなく、定常音に近い音に聞こえるようなモデル騒音を作成することにする。これは収録した実際の騒音を騒音により異なるが短い区間で区切り、それを少しずつずらしながら重ねていくことで作成する。このモ

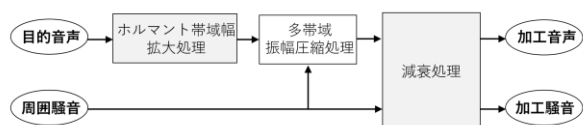


図5 聴覚機能低下模擬の処理過程

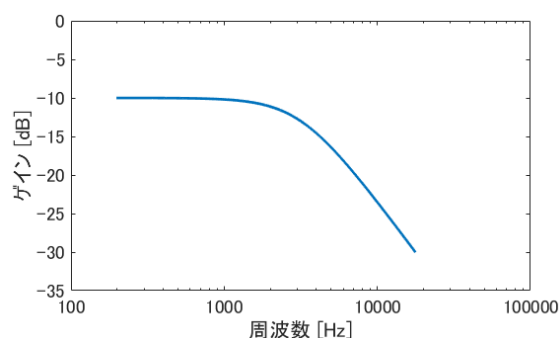


図6 減衰処理フィルターの周波数特性

表1 モデル騒音の種類と S/N

モデル騒音 番号	元となった騒音の内容	騒音の呈示 音圧(dB SPL)	S/N(dB)
1	電車の発車音1	72	+1
2	電車のホーン	72	+1
3	特急列車の通過音	75	-2
4	雑踏(マルチトーカーノイズ)	78	-5
5	電車の停車音1	76	-3
6	電車の停車音2	77	-4
7	電車の停車中の音	81	-8
8	電車の発車音2	79	-6
9	電車の発車音3	73	±0
10	電車の発車音4	72	+1
11	雨音	73	±0
12	足音	71	+2
13	電車の通過音	70	+3
14	踏切	65	+8
15	貨物列車の発車音	73	±0
16	駅構内の工事の音	77	-4
17	飛行機の通過音	78	-5
18	車のクラクション	75	-2
19	バイクの通過音	76	-3
20	鳥の鳴き声	70	+3

デル騒音は20種類作成した。その内容を表1に示す。

c) 音声と騒音の呈示音圧レベル

音声の呈示音圧レベルは、聴取最適音圧レベル(MCL)が一般的に約60dB(A)(音声であればおよそ63dB SPL)であることから、駅構内を想定している為それより少し大きめの呈示音圧レベルとして、73dB SPLとする。ちなみに、本研究は、騒音下での文章理解度がやや悪い時、つまり騒音下で文章がやや聞き取りにくい時に、聞き取りやすくなるよう加工するシステムの開発を目指している。そこで、一般に単語理解度が90%のとき、文章理解度は理想伝送系で98%となる⁹⁾ことが示されていることを考慮して、実験を行う際は、単語理解度が約70～75%となるようなS/Nで音声を呈示することとする。そこで、予備実験として各騒音下での単語の書き取り実験を行う。被験者は、純音聴力検査の結果、聴覚機能に異常がなく、かつ被験者となることに同意を得られた20～25歳の男女15名である。この予備実験の結果を表1に示す。

(3) 教師データ収集

被験者は単語を知らない、つまり予備実験を行っていない20～25歳の男女20名である。この被験者も、純音聴力検査の結果、聴力に異常がなく、かつ被験者となることに同意を得ている。ここで、被験者1人につき、呈示される騒音は4種類のモデル騒音とし、その騒音下での圧縮対象となる目的音声は、騒音1種類につき、25種類の単語とする。呈示される目的音声は各モデル騒音下での正答率が約70～75%となるようなS/Nで呈示している為、この場合、未圧縮状態ではこれらの25種類の単語を全て正確に聞き取ることはできない。そこで、被験者に“呈示される単語を、全て正答するまで実験用ア



図7 教師データ作成実験における被験者操作画面

アプリケーションプログラムを繰り返し操作せよ“との指示を与え、被験者毎に各モデル騒音下での適切な圧縮レベルを決定させた。

ここで図7に、この実験のために制作した、目的音声の各帯域における圧縮レベルを変化させることができるアプリケーションプログラムのインターフェースを示す。赤枠の中の5組の増減ボタンは、帯域分けした目的音声の各帯域と対応しており、このボタンを操作することで圧縮レベルを自由に変更することができる。ある一組の圧縮レベルで全ての単語を正答することができるプログラムは終了し、その時の各帯域の圧縮レベルを保存する。この記録したデータが教師データの候補となる。

(4) 騒音毎の最適な教師データ選定

収集したデータは、騒音毎に4名の被験者の圧縮レベル、つまり騒音1種類に対して4組の圧縮レベルが存在する。そこで、どの圧縮レベルで圧縮した音声聞き取りやすいかを調べるために、この4組の圧縮レベルでそれぞれ圧縮した音声を騒音下で呈示し、一対比較法を用いて実験を行う。

被験者は、予備実験または教師データ作成実験を行った内の9名で行った。ここで、音を1対1で比較する際に、同時に聞かせることはできないため呈示順序を考える必要がある。1つの騒音において、圧縮レベルは4組存在することから、2組ずつ比較するための並べ方は12通りある。今回はこれらを2回ずつ、つまり24回の1対1比較を、全20種類のモデル騒音下において行わせる。このようにして最も選択された音声の圧縮レベルを、その騒音の最適な圧縮レベルとして選定した。この実験を行うことで、1種類の騒音に対して4組の圧縮レベルが存在していたものを、騒音と適切な圧縮レベルを対応させることで、これを教師データとすることにした。

6. ニューラルネットワークによる学習

今回使用するニューラルネットワークは、入力層5素子、中間層3素子、出力層5素子であり、バックプロパゲーション法を用いて予め学習させたものである。トレーニングデータは20種類のモデル騒音を5帯域に帯

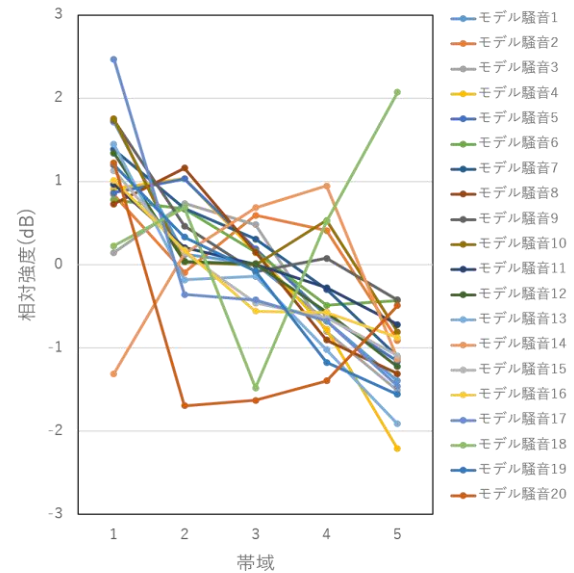


図8 加工後のモデル騒音の分析結果

域分けした各帯域の音圧レベル、教師データは先の教師データ作成実験で求めた騒音毎の適切な圧縮レベルとする。

ここで、各モデル騒音の音圧レベルは、巨視的に見ると互いに類似していて、これらのデータをそのまま学習に用いると学習の収束が損なわれる可能性があると考えられる。そこで、実際にニューラルネットワークに学習させる時には、モデル騒音の音圧レベル各々に加工を施す。この加工は、モデル騒音の音圧レベル x_i に対し、加工後を X_i とすると、

$$X_i = \frac{x_i - 66}{9} \quad (2)$$

で求められる。加工後の分析結果を図8に示す。これらのデータを学習させることで、実際の周囲騒音の音圧レベルを入力した時に、その騒音下での適切な圧縮レベルを出力することができるニューラルネットワークになることが期待できる。

7. 評価実験

(1) 実験の概要

本研究では、高齢者にとっての了解度を向上させるような音声加工処理システムの実現を目指している。しかし、教師データ作成実験等においては被験者の負担等の問題から、高齢者模擬を施した音声を用いて被験者は若年者としていた。そこで、高齢者を被験者として試作したシステムの評価実験を行う。騒音下で目的音声に多帯域振幅圧縮処理を施し、目的音声の受聴明瞭度がどの程度改善されたかを検証するために、単語書き取り実験と主観評価実験を交互に行う。被験者は、被験者となることに同意を得ている、71歳の男性（被験者A）と81歳

の女性（被験者 B）とする。

（２）実験方法

a) 単語書き取り実験

この実験は、ある騒音下で未圧縮音声と実験システムで圧縮を施した音声を、等しい S/N で呈示し、その内容を書き取らせる実験である。そして、未圧縮音声と圧縮音声の正答率を比較して、各騒音下での単語理解度がどの程度改善されたかを評価値とする。

実験に使用する呈示音声は、教師データー作成実験で使用した単語と同様に、補聴器適合評価用 CD の“TY-89”に収録されている 2 音節の単語 50 種類と 3 音節の単語 50 種類、計 100 種類とする。この音声には高齢者の聴覚機能低下を模擬する加工はしない。また、実験に使用する騒音は、教師データー作成時に使用したモデル騒音 3 種類（モデル騒音 5, 13, 14）と、ピンクノイズを用いて実際の騒音のスペクトル構造を模擬したモデル騒音 2 種類（ここで、モデル騒音 21, 22 とする）、計 5 種類の騒音を用いた。これらの呈示音圧レベルを表 2 に示す。

ここで、被験者 1 人につき、呈示される騒音は 5 種類のモデル騒音とし、その騒音下での呈示される音声は、騒音 1 種類につき、40 種類の単語とする。これらを未圧縮音声と圧縮音声で比較をするために 2 つに分ける。今回は被験者の数が少ないため、さらにそれを 2 つに分けることにする。つまり、実験は 1 種類の騒音につき未圧縮音声と圧縮音声をそれぞれ 2 回（計 4 回）行い、1 回の実験で呈示される単語は 10 種類とする、1 回分の書き取り実験が終了した後、後述する主観評価実験を行う。

b) 主観評価実験

この実験は、先の単語書き取り実験で 1 回につき 10 種類の単語の書き取りが終了した後に、それらの単語がどう聞こえたか、その印象について分量評定してもらうものである。印象については、以下の 4 項目について問うことにする。

- ・ 音声の聞き取りやすさ
- ・ 音声の煩わしさ
- ・ 騒音に対しての音声の大きさ
- ・ 総合的な良さ

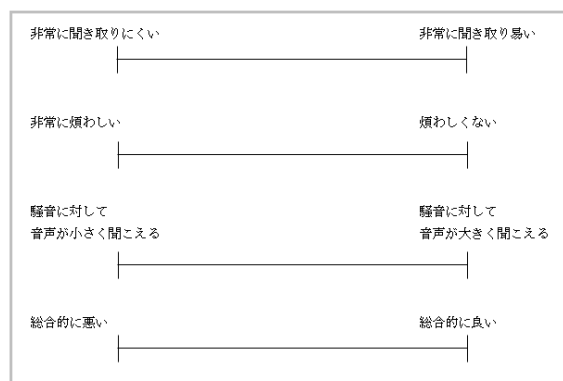
そして、被験者の評価において、評価が悪い方が“0”，良い方が“100”としてシステムの評価を行う。この実験で使用した、被験者用の検査用紙を図 9 に示す。

表 2 評価実験におけるモデル騒音の種類と S/N

モデル騒音番号	元となった騒音の内容	騒音の呈示音圧 (dB SPL)	S/N (dB)
5	電車の停車音1	76	-3
13	電車の通過音	70	+3
14	踏切	65	+8
21	踏切2	65	+8
22	電車のホーン	72	+1

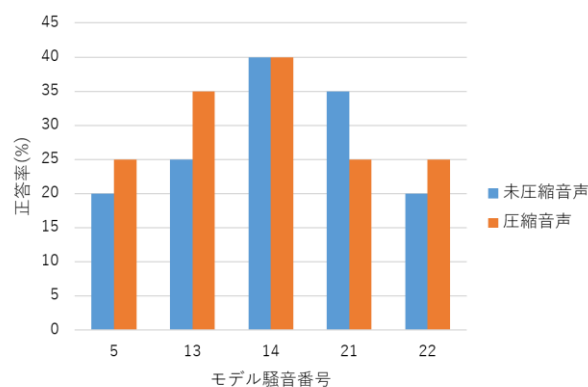
（３）実験結果

単語書き取り実験の結果として、図 10 の(a)と(b)にそれぞれ被験者 A と被験者 B の正答率と、図 11 に騒音毎の改善された単語理解度の平均を示す。この実験では、騒音毎に結果のばらつきがあり、圧縮をすることで改善できていない騒音もあるが、最大で約 12% の単語理解度の改善が見られた。また、主観評価に関しては、有意な差が見られたものはなかった。

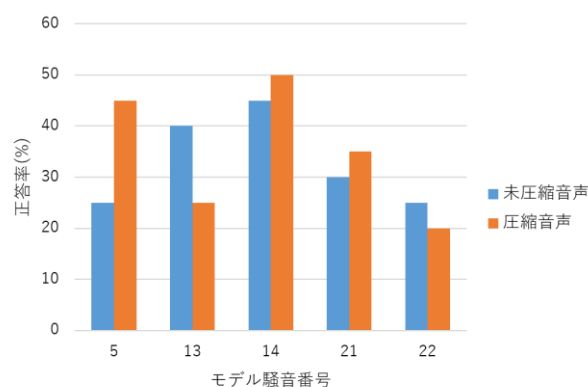


評価項目	左側（悪い）	右側（良い）
聞き取りやすさ	非常に聞き取りにくい	非常に聞き取り易い
煩わしさ	非常に煩わしい	煩わしくない
音声の大きさ	騒音に対して音声小さく聞こえる	騒音に対して音声大きく聞こえる
総合的な良さ	総合的に悪い	総合的に良い

図 9 主観評価実験で用いる検査用紙



(a) 被験者 A



(b) 被験者 B

図 10 騒音下での単語理解度

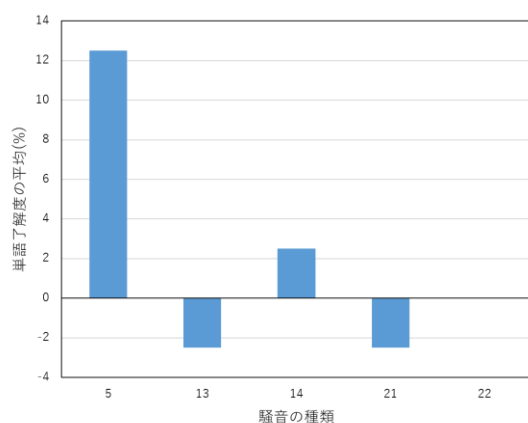


図 1 1 騒音毎の改善された単語理解度の平均

8. 考察と今後の研究課題

評価実験の結果、今回の実験は被験者が少ないため統計的に断言することはできないが、高齢者の騒音下における音声聴取の改善は、平均値は大きいものの、統計的に有意なものではなかった。これは、本システムで帯域分けを行う時に使用しているフィルターの次数が低いことが原因として挙げられる。現在は先行研究に倣って次数を 91 次としているが、サンプリング周波数を先行

研究の 11025Hz から 44100Hz に変更したため、フィルターの次数を上げてシステム全体を作り直す必要がある。

さらに、実時間処理をする上で、処理をしたあとに拡声器から出力された音声は、周囲騒音の一部として分析対象に含まれてしまうことの補正を加える必要がある。

参考文献

- 1) 鈴木郁，越智寛毅：高齢者にとっての了解度改善を目的とした、拡声器のための音声加工処理，人間工学会誌，Vol. 40，No. 6，pp. 289-301，2004
- 2) 上田祐市，阿川貴之，青山正純，渡邊亮：合成フィルタによるデジタル補聴器のための振幅圧縮処理，電子情報通信学会論文誌，Vol.J81-A，No.12，pp.1728-1738，1998
- 3) 田中美郷他：補聴器適合評価機器の試作に関する研究，文部省科学研究費研究成果報告書，1989
- 4) 鈴木郁，林善男：雑音下における無意味後の若年者および高齢者による聴取について，人間工学会誌，Vol.27，No.1，pp25-33，1991
- 5) 日本音響学会，音響用語辞典，コロナ社，526，1988